# Bayesian Analysis for Change－Point Linear Regression <br> Models＊ 

Tang Yincai ${ }^{\star}$ Wang Pingping<br>（School of Finance and Statistics，East China Normal University，Shanghai，200241）<br>Chen Hui<br>（Shanghai South West Weiyu Middle School，Shanghai，200233）


#### Abstract

This article considers Bayesian inference of the linear regression model with one change point in observations，provided that the prior distribution of the change point is the beta－binomial distri－ bution or the power prior introduced by Ibrahim et al．（2003）and the variances of the observations on two sides of the change point are the same．We get closed forms of the posterior distributions of the change point，the regression coefficients and the common variance．This not only generalizes the result of Ferreira（1975）from the the discrete uniform prior distribution of the change point $t$ to the beta－binomial distribution which can well describe the shape of the change point distri－ bution，but also can be further generalized to the power prior distribution of the change point， which included the historical information．Simulation shows higher performance or accuracy of the Bayesian method when the change point follows the beta－binomial and power prior．


Keywords：Beta－binomial prior，power prior，change point，Bayesian estimation，linear model．

AMS Subject Classification：62J05．

## §1．Introduction

The change point is a specific parameter that is introduced to account for the abrupt change in a non－smooth manner at a particular point in time．Typically，there exist two types of change points．The first one is in change of the location of independent variables of the observations．The second one is in the change of the observations themselves．The main difference of them is that in the first change－point problem the independent variables are in the natural order of time，while in the second model the indices of the observations are the time scales．In other words，the first change－point model consists of two segments

[^0]while the second is a sort of mixture of two models．Take linear regression model with only one change point for example．Let $\left(x_{1}, y_{1}\right),\left(x_{1}, y_{2}\right), \ldots,\left(x_{n}, y_{n}\right)$ be the observations of size $n$ ．The first kind of change－point simple linear regression model with one change point can be expressed as
\[

y_{(i)}= $$
\begin{cases}a_{1} x_{(i)}+b_{1}+\varepsilon_{1}, & 1 \leq i \leq t \\ a_{2} x_{(i)}+b_{2}+\varepsilon_{2}, & t<i \leq n\end{cases}
$$
\]

where $a_{1} \neq a_{2}$ and $\left(x_{(t)}, y_{(t)}\right)$ or simply $t$ is the change point which is the cut point for the independent variables．The second type has the form as follows．

$$
y_{i}= \begin{cases}a_{1} x_{i}+b_{1}+\varepsilon_{1}, & 1 \leq i \leq t \\ a_{2} x_{i}+b_{2}+\varepsilon_{2}, & t<i \leq n\end{cases}
$$

where $a_{1} \neq a_{2}$ and $\left(x_{t}, y_{t}\right)$ or simply $t$ is the change point which is the cut point for the observations．In both types of change－point problems，the distributions of $\varepsilon_{1}$ and $\varepsilon_{2}$ can be the same or not．

The change－point problem occurs frequently in medical research，product life time degradation experiment and time series data．For example，cancer incidence rates remain relatively stable for people at a younger age，but change drastically after a certain age threshold（MacNeill and Mao，1995）．The data obtained from a group of preschool boys indicates that their weight／height ratio relates to their age in one way before a certain age but that the functional relation of the two changes afterwards（Gallant，1977）．Example arises from a study of the risk of myocardial infarction，which showed a sharp decrease in risk at low alcohol intakes and a dramatic increase after reaching a certain a mount of daily alcohol consumption（Pastor and Guallar，1988）．These three examples are the first type of change－point regression models．Another example is that two or more testers get samples from one experiments and the differences of testers may lead to the different results of the experiment．This may lead to the second type of the change－point problem． In this paper we only discuss the second type of change－point problem in the regression setting and we mean the second one in the later discussion if not otherwise stated．

From the perspective of the development of statistics，methods about change－point problems are not mature．For specific problems，there have been some particular effective methods including Bayesian method，Schwarz information criterion method，maximum likelihood method，nonparametric method and the least square method et al．Ferreira （1975）estimated the parameters of linear regression model with one change point using the Bayesian method．Chin Choy and Broemeling（1980）dealt with the problem of generalized linear regression model with one change point．

Ferreira（1975）used the discrete uniform distribution as the prior distribution of the change point．We generalize the result of Ferreira（1975）from the discrete uniform prior to the beta－binomial prior and power prior，which helps elicit useful historical data or prior information in practice for the change point．Beta－binomial distribution，a compound distribution of the binomial distribution binom $(n, p)$ with respect to the beta distribution $\operatorname{beta}(a, b)$ for the success probability $p$ ，is the one dimension discrete random variable distribution taking integral values from 0 to $n$ and resembles the the shape of the beta distribution which takes continuous values from 0 to 1 ，though．Power prior（Ibrahim et al．，2003）is an informative prior that incorporate the historical data into the current study by quantifying it with a suitable prior distribution on the model parameter．

The structure of the paper is as follows．In Section 1，we introduce the change－ point problem for linear regression model and Section 2 gives the assumptions needed for the Beysian inference．Section 3 gives the technique development of Bayesian analysis of linear regression model with one change point and presents the closed forms of the marginal posterior distribution of the parameters．Section 4 is a numerical example which shows the effectiveness of the Bayesian estimation．

## §2．Bayesian Analysis

## 2．1 Assumptions

Assume that $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right), \ldots,\left(x_{n}, y_{n}\right)$ are $n$ pairs of observations which follow the linear regression model with change point $t$ ．That is，

$$
y_{i}= \begin{cases}a_{1}\left(x_{i}-\bar{x}_{1}\right)+b_{1}+\varepsilon_{i}, \quad \operatorname{Var}\left(y_{i}\right)=\sigma^{2}, & i=1, \ldots, t ;  \tag{2.1}\\ a_{2}\left(x_{i}-\bar{x}_{2}\right)+b_{2}+\varepsilon_{i}, \operatorname{Var}\left(y_{i}\right)=\sigma^{2}, & i=t+1, \ldots, n,\end{cases}
$$

where

$$
\bar{x}_{1}=\frac{1}{t} \sum_{i=1}^{t} x_{i}, \quad \bar{y}_{1}=\frac{1}{t} \sum_{i=1}^{t} y_{i}, \quad \bar{x}_{2}=\frac{1}{n-t} \sum_{i=t+1}^{n} x_{i}, \quad \bar{y}_{2}=\frac{1}{n-t} \sum_{i=t+1}^{n} y_{i} .
$$

To establish the linear regression model，four points are needed at least and $t=2,3, \ldots, n-$ 2.

The regression coefficients in the first part of the model are not exactly the same as the second part of the model．In addition，we assume for simplicity that $\operatorname{Var}\left(y_{i}\right)=\sigma^{2}, i=$ $1, \ldots, n$ ．The main purpose of the paper is the estimation of the unknown parameters $t, a_{1}, b_{1}, a_{2}, b_{2}$ and $\sigma^{2}$ ．

We further assume that
1．all prior distributions of the parameters are independent．

2．the prior distributions of the regression coefficients $a_{1}, b_{1}, a_{2}$ and $b_{2}$ in the quadratic linear regression model（2．1）follows the flat distribution on $(-\infty, \infty)$ ．

3．the prior distribution of the variance $\sigma^{2}$ follows the conjugate inverse gamma dis－ tribution $\operatorname{IG}(\alpha, \lambda)$ with pdf

$$
\begin{equation*}
\pi\left(\sigma^{2}\right)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\left(\frac{1}{\sigma^{2}}\right)^{\alpha+1} \exp \left\{-\frac{\lambda}{\sigma^{2}}\right\} \tag{2.2}
\end{equation*}
$$

where $\lambda$ and $\alpha$ are hyper－parameters and $\Gamma(\alpha)$ is the gamma function，$\Gamma(\alpha)=\int_{0}^{\infty} t^{\alpha-1} \mathrm{e}^{-t} \mathrm{~d} t$ ． Notice that the inverse gamma distribution with $\alpha=0$ and $\lambda=0$ is a noninformative prior for $\sigma^{2}$ ；that is $\pi\left(\sigma^{2}\right) \propto 1 / \sigma^{2}$ ．

4．two kinds of prior distribution of the change point $t$ are considered here．The first one is beta－binomial，beta－bin $(n, a, b)$ ，whose probability mass function（pmf）is defined as

$$
\begin{equation*}
\pi(t \mid n, a, b)=\binom{n}{t} \frac{B(t+a, n-t+b)}{B(a, b)}, \quad t=0,1, \ldots, n \tag{2.3}
\end{equation*}
$$

where $n$ is the sample size，$a$ and $b$ are the hyper－parameters，and $B(x, y)$ is the beta
 function defined as $B(x, y)=\int_{0}^{1} t^{x-1}(1-t)^{y-1} \mathrm{~d} t$ ．

Beta－binomial distribution，also called negative hypergeometric or Polya distribution， is the compound distribution of the beta and binomial distributions．In the Bayesian setting，it is obtained as the marginal distribution of the number of successes in $n$ Bernoulli trials which follows a binomial distribution，for the success proportion or the probability of success when it takes the conjugate beta prior $\operatorname{Beta}(a, b)$ ．A beta－binomial distribution returns a integer value between 0 and $n$ ．It can be looked upon as the discrete counterpart of the beta distribution which returns a continuous value between 0 and 1 ．See Figure 1 for the comparison of the beta and beta－binomial distributions for different combinations of $a$ and $b$ given $n=20$ ．As mentioned earlier，for the change－point problem we discuss，$t$ only takes values of $2,3, \ldots, n-2$ ．

Another prior distribution of change point $t$ is the power prior．Denote $D(n, y, X)$ as the current study data，where $n$ denotes the sample size，$y$ denotes the $n \times 1$ response vector，and $X$ the $n \times p$ covariates matrix．Similarly，the historical data is denoted by $D_{0}\left(n_{0}, y_{0}, X_{0}\right)$ ．Further，the marginal likelihood of the change point $t$ which can be got by integrating the joint likelihood with respect to other parameters is denoted by $L(t \mid D)$ ． Let $\pi_{0}(t)$ denote the prior distribution before the historical data，in this paper $\pi_{0}(t) \propto 1$ ． Given $a_{0}$ ，we define the power prior distribution of $t$ for the current study as

$$
\begin{equation*}
\pi\left(t \mid D_{0}, a_{0}\right) \propto L\left(t \mid D_{0}\right)^{a_{0}} \pi_{0}(t) \tag{2.4}
\end{equation*}
$$

where $a_{0}$ is a scalar parameter that weights the historical data relative to the likelihood for the current study．


Figure 1 Comparison of beta and beta－binomial distributions

## 2．2 Model Formulation

According to Bayesian formulation，the likelihood function of the linear regression model（2．1）with one change point is as follows，

$$
\begin{aligned}
& L(\boldsymbol{\theta} \mid \boldsymbol{\tau}) \\
= & \left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{n} \exp \left\{-\frac{1}{2 \sigma^{2}}\left[\sum_{i=1}^{t}\left(y_{i}-a_{1}\left(x_{i}-\bar{x}_{1}\right)-b_{1}\right)^{2}+\sum_{i=t+1}^{n}\left(y_{i}-a_{2}\left(x_{i}-\bar{x}_{2}\right)^{2}-b_{2}\right)^{2}\right]\right\} \\
= & \left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{n} \exp \left\{-\frac{1}{2 \sigma^{2}}\left[\sum_{i=1}^{t}\left(\left(y_{i}-\bar{y}_{1}\right)-a_{1}\left(x_{i}-\bar{x}_{1}\right)-\left(b_{1}-\bar{y}_{1}\right)\right)^{2}\right.\right. \\
& \left.\left.+\sum_{i=t+1}^{n}\left(\left(y_{i}-\bar{y}_{2}\right)-a_{2}\left(x_{i}-\bar{x}_{2}\right)-\left(b_{2}-\bar{y}_{2}\right)\right)^{2}\right]\right\}
\end{aligned}
$$

where $\boldsymbol{\theta}=\left(a_{1}, a_{2}, b_{1}, b_{2}, \sigma^{2}, t\right), \boldsymbol{\tau}=\left\{\left(x_{i}, y_{i}\right), i=1,2, \ldots, n\right\}, t=2,3, \ldots, n-2$ ．

For the sake of simplicity，let

$$
\begin{aligned}
& T_{1}^{2}=\sum_{i=1}^{t}\left(\left(y_{i}-\bar{y}_{1}\right)-a_{1}\left(x_{i}-\bar{x}_{1}\right)-\left(b_{1}-\bar{y}_{1}\right)\right)^{2} \\
& T_{2}^{2}=\sum_{i=t+1}^{n}\left(\left(y_{i}-\bar{y}_{2}\right)-a_{2}\left(x_{i}-\bar{x}_{2}\right)-\left(b_{2}-\bar{y}_{2}\right)\right)^{2}
\end{aligned}
$$

Then the likelihood function can be written as

$$
L(\boldsymbol{\theta} \mid \boldsymbol{\tau})=\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{n} \exp \left\{-\frac{1}{2 \sigma^{2}}\left(T_{1}^{2}+T_{2}^{2}\right)\right\}
$$

Thus，given the priors of $\boldsymbol{\theta}=\left(a_{1}, a_{2}, b_{1}, b_{2}, \sigma^{2}, t\right)$ above，we get the joint posterior distri－ bution of $\boldsymbol{\theta}$

$$
\begin{align*}
\pi(\boldsymbol{\theta} \mid \boldsymbol{\tau}) & \propto L(\boldsymbol{\theta} \mid \boldsymbol{\tau}) \pi(\boldsymbol{\theta}) \\
& \propto\left(\frac{1}{\sigma^{2}}\right)^{(n+2) / 2} \exp \left\{-\frac{T_{1}^{2}+T_{2}^{2}}{2 \sigma^{2}}\right\} \pi(t) \tag{2.5}
\end{align*}
$$

where $\pi(t)$ is the prior distribution of the change point $t$ ，which can be either the beta－ binomial prior（2．3）or the power prior（2．4）．

## $\S 3 . \quad$ Posteriors

For the convenience of the development of the posterior distributions of the regression coefficients and the change point，we need to decompose $T_{1}^{2}$ and $T_{2}^{2}$ with the help of the notations below

$$
\begin{array}{ll}
S_{x 2}^{(1)}=\sum_{i=1}^{t}\left(x_{i}-\bar{x}_{1}\right)^{2}, & S_{x 2}^{(2)}=\sum_{i=t+1}^{n}\left(x_{i}-\bar{x}_{2}\right)^{2} \\
S_{y 2}^{(1)}=\sum_{i=1}^{t}\left(y_{i}-\bar{y}_{1}\right)^{2}, & S_{y 2}^{(2)}=\sum_{i=t+1}^{n}\left(y_{i}-\bar{y}_{2}\right)^{2} \\
S_{x y}^{(1)}=\sum_{i=1}^{t}\left(x_{i}-\bar{x}_{1}\right)\left(y_{i}-\bar{y}_{1}\right), & S_{x y}^{(2)}=\sum_{i=t+1}^{n}\left(x_{i}-\bar{x}_{2}\right)\left(y_{i}-\bar{y}_{2}\right), \\
S_{1}^{2}=S_{y 2}^{(1)}-S_{x y}^{(1)}{ }^{2} / S_{x 2}^{(1)}, & S_{2}^{2}=S_{y 2}^{(2)}-S_{x y}^{(2)^{2}} / S_{x 2}^{(2)} \\
S^{2}=S_{1}^{2}+S_{2}^{2} &
\end{array}
$$

Then $T_{1}^{2}$ and $T_{2}^{2}$ can be decomposed into three terms，

$$
\begin{aligned}
& T_{1}^{2}=S_{1}^{2}+t\left(b_{1}-\bar{y}_{1}\right)^{2}+S_{x 2}^{(1)}\left(a_{1}-S_{x y}^{(1)} / S_{x 2}^{(1)}\right)^{2} \\
& T_{2}^{2}=S_{2}^{2}+(n-t)\left(b_{2}-\bar{y}_{2}\right)^{2}+S_{x 2}^{(2)}\left(a_{2}-S_{x y}^{(2)} / S_{x 2}^{(2)}\right)^{2}
\end{aligned}
$$

## 3．1 Marginal Posterior Distribution of $t$

According to the properties of $\chi^{2}$ and Student＇s $t$ distribution，integrating the join－ t posterior distribution function about $\sigma^{2}, b_{1}, a_{1}, b_{2}$ and $a_{2}$ respectively，we can get the marginal posterior distribution of $t$ ．

First，integrate（2．5）with respect to $\sigma^{2}$ ，which can be regarded as an integration of the density function the inverse gamma distribution with parameters $\lambda=\left(T_{1}^{2}+T_{2}^{2}\right) / 2$ and $\alpha=n / 2$ ．Then it gives

$$
\begin{align*}
& \int_{0}^{\infty}\left(\frac{1}{\sigma^{2}}\right)^{(n+2) / 2} \exp \left\{-\frac{T_{1}^{2}+T_{2}^{2}}{2 \sigma^{2}}\right\} \pi(t) \mathrm{d} \sigma^{2} \\
= & \frac{\Gamma(n / 2)}{\left(\left(T_{1}^{2}+T_{2}^{2}\right) / 2\right)^{n / 2}} \pi(t) \propto\left(T_{1}^{2}+T_{2}^{2}\right)^{-n / 2} \pi(t) . \tag{3.1}
\end{align*}
$$

Secondly，integrate（3．1）with respect to $b_{1}$ ．With

$$
B=S_{1}^{2}+S_{x 2}^{(1)}\left(a_{1}-S_{x y}^{(1)} / S_{x 2}^{(1)}\right)^{2}+T_{2}^{2},
$$

we have

$$
\begin{align*}
& \int_{-\infty}^{\infty}\left(T_{1}^{2}+T_{2}^{2}\right)^{-(n+2) / 2} \pi(t) \mathrm{d} b_{1}=\int_{-\infty}^{\infty}\left[t\left(b_{1}-\bar{y}_{1}\right)^{2}+B\right]^{-n / 2} \pi(t) \mathrm{d} b_{1} \\
= & B^{-n / 2} \int_{-\infty}^{\infty}\left[1+\frac{t\left(b_{1}-\bar{y}_{1}\right)^{2}}{B}\right]^{-n / 2} \pi(t) \mathrm{d} b_{1} \\
= & B^{-n / 2} \int_{-\infty}^{\infty}\left[1+\frac{\left(\sqrt{(n-1) t / B}\left(b_{1}-\bar{y}_{1}\right)\right)^{2}}{n-1}\right]^{-n / 2} \pi(t) \mathrm{d} b_{1} \\
\propto & t^{-1 / 2} B^{-(n-1) / 2} \pi(t) . \tag{3.2}
\end{align*}
$$

Let

$$
A=S_{1}^{2}+T_{2}^{2}
$$

Integrating（3．2）with respect to $a_{1}$ gives rise to

$$
\begin{equation*}
\int_{-\infty}^{\infty} t^{-1 / 2} B^{-(n-1) / 2} \pi(t) \mathrm{d} a_{1} \propto\left(t S_{x 2}^{(1)}\right)^{-1 / 2} A^{-(n-2) / 2} \pi(t) . \tag{3.3}
\end{equation*}
$$

Similarly，integrating（3．3）with respect to $b_{2}$ and $a_{2}$ respectively，we get the posterior distribution of $t$ ，

$$
\pi(t \mid \boldsymbol{\tau}) \propto\left(t(n-t) S_{x 2}^{(1)} S_{x 2}^{(2)}\right)^{-1 / 2}\left(S^{2}\right)^{-(n-4) / 2} \pi(t), \quad t=2,3, \ldots, n-2
$$

## 3．2 Marginal Posterior Distributions of $a_{1}$ and $a_{2}$

Similar to the process of getting the marginal posterior distribution of $t$ ，integrating （2．5）with respect to $\sigma^{2}, b_{1}, a_{1}$ and $b_{2}$ in turn and adding all values up with respect to $t$ ，
we can get the posteriors of $a_{2}$ ．

$$
\pi\left(a_{2} \mid \boldsymbol{\tau}\right) \propto \sum_{t=2}^{n-2}\left(1+\frac{\left(a_{2}-S_{x y}^{(2)} / S_{x 2}^{(2)}\right)^{2}}{S^{2} / S_{x 2}^{(2)}}\right)^{-(n-3) / 2} \frac{\pi(t \mid \boldsymbol{\tau})}{\left(S^{2} / S_{x 2}^{(2)}\right)^{1 / 2}}
$$

Similarly，we get the marginal posterior distribution of $a_{1}$ as follows，

$$
\pi\left(a_{1} \mid \boldsymbol{\tau}\right) \propto \sum_{t=2}^{n-2}\left(1+\frac{\left(a_{1}-S_{x y}^{(1)} / S_{x 2}^{(1)}\right)^{2}}{S^{2} / S_{x 2}^{(1)}}\right)^{-(n-3) / 2} \frac{\pi(t \mid \boldsymbol{\tau})}{\left(S^{2} / S_{x 2}^{(1)}\right)^{1 / 2}}
$$

Thus，the marginal posterior distribution of $a_{j}(j=1,2)$ is a weighted non－standar－ dized Student＇s $t$ distribution with location $S_{x y}^{(j)} / S_{x 2}^{(j)}$ ，scale $S^{2} / S_{x 2}^{(j)}$ and degree of freedom $n-4$ ．The weights are $\pi(t \mid \boldsymbol{\tau}), t=2,3, \ldots, n-2$ ，the posterior probability mass function values．

## 3．3 Marginal Posterior Distributions of $b_{1}$ and $b_{2}$

In a similar manner，integrating（2．5）with respect to $\sigma^{2}, a_{1}, b_{1}$ and $a_{2}$ in turn and adding all values up with respect to $t$ ，we can get the marginal posterior distribution of $b_{2}$ is

$$
\pi\left(b_{2} \mid \boldsymbol{\tau}\right) \propto \sum_{t=2}^{n-2}\left(1+\frac{\left(b_{2}-\bar{y}_{2}\right)^{2}}{S^{2} /(n-t)}\right)^{-(n-3) / 2} \frac{\pi(t \mid \boldsymbol{\tau})}{\left(S^{2} /(n-t)\right)^{1 / 2}}
$$

And the marginal posterior distribution of $b_{1}$ is

$$
\pi\left(b_{1} \mid \boldsymbol{\tau}\right) \propto \sum_{t=2}^{n-2}\left(1+\frac{\left(b_{1}-\bar{y}_{1}\right)^{2}}{S^{2} / t}\right)^{-(n-3) / 2} \frac{\pi(t \mid \boldsymbol{\tau})}{\left(S^{2} / t\right)^{1 / 2}} .
$$

Thus we see that the marginal posterior distribution of $b_{j}(j=1,2)$ is a weighted non－ standardized Student＇s $t$ distribution with location $\bar{y}_{j}$ ，scale $S^{2} / t$ and degree of freedom $n-4$ ．The weights are values of the posterior probability mass function $\pi(t \mid \boldsymbol{\tau})$ ．

## 3．4 Marginal Posterior Distribution of $\sigma^{2}$

Integrating（2．5）with respect to $b_{1}, a_{1}, b_{2}$ and $a_{2}$ respectively according to the property of normal distribution and adding up all values of $t$ ，we can get the marginal posterior distribution $\sigma^{2}$ ．

$$
\pi\left(\sigma^{2} \mid \boldsymbol{\tau}\right) \propto \sum_{t=2}^{n-2}\left(\frac{S^{2}}{\sigma^{2}}\right)^{(n-6) / 2} \mathrm{e}^{-S^{2} /\left(2 \sigma^{2}\right)} S^{2} \pi(t \mid \boldsymbol{\tau})
$$

Therefore，the marginal posterior distribution of $\sigma^{2}$ has the nice property that it is a mixture of inverse gamma distributions with shape parameter $(n-8) / 2$ and rate parameter $S^{2} / 2$ ．The weights are still the values of $\pi(t \mid \boldsymbol{\tau})$ ．

## §4．Numerical Example

Here，we analysis the data given in Quandt（1958）and Ferreira（1975）using Bayesian Method．These data were created accordingly to the model

$$
y_{i}= \begin{cases}0.7\left(x_{i}-\bar{x}_{1}\right)+9.15+\varepsilon_{i}, & i=1, \ldots, 12 \\ 0.5\left(x_{i}-\bar{x}_{2}\right)+11+\varepsilon_{i}, & i=13, \ldots, 20\end{cases}
$$

where the error were generated following a $N(0,1)$ distribution．
And here we using five prior distributions of change point．The first three were used by Ferreira（1975）and the last two are beta－binomial distribution and the power prior which are introduced in this paper．Figure 2 shows their pmf．

$$
\begin{aligned}
& \pi_{1}(t) \propto 1 \\
& \pi_{2}(t) \propto(t(n-t))^{1 / 2} \\
& \pi_{3}(t) \propto\left(t(n-t) S_{x 2}^{(1)} S_{x 2}^{(2)}\right)^{1 / 2} \\
& \pi_{4}(t) \propto\binom{n}{t} \frac{B(t+8, n-t+5)}{B(8,5)} \\
& \pi_{5}(t) \propto\left(t_{0}\left(n_{0}-t_{0}\right) S_{x 20}^{(1)} S_{x 20}^{(2)}\right)^{1 / 2}
\end{aligned}
$$

The deviation and mean squared errors of the estimates computed with 1000 repetitions are listed in Table 1．Note that the MSE under $\pi_{3}$ is smaller than others．And the marginal distribution of five priors are given in Figures 3－7．

Table 1 Bayesian estimates corresponding to five prior distributions

|  | prior | $t$ | $a_{1}$ | $a_{2}$ | $b_{1}$ | $b_{2}$ | $\sigma^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Deviation | $\pi_{1}$ | －0．10483 | －0．0027229 | －0．0016050 | 0.033114 | －0．10193 | 0.11909 |
|  | $\pi_{2}$ | －0．12949 | －0．0016373 | －0．0009565 | 0.021626 | －0．09612 | 0.11259 |
|  | $\pi_{3}$ | －0．16703 | 0.0006213 | －0．0036355 | 0.005348 | －0．08512 | 0.09940 |
|  | $\pi_{4}$ | 0.02223 | －0．0035617 | －0．0098051 | 0.040395 | －0．04365 | 0.09344 |
|  | $\pi_{5}$ | －0．03395 | －0．0034710 | －0．0028541 | 0.012081 | －0．06851 | 0.10158 |
| MSE | $\pi_{1}$ | 4.64793 | 0.7167612 | 8.0472600 | 0.328598 | 0.31405 | 0.24564 |
|  | $\pi_{2}$ | 4.11569 | 0.5715159 | 7.1300023 | 0.295608 | 0.28802 | 0.23424 |
|  | $\pi_{3}$ | 2.68788 | 0.0139487 | 0.0370354 | 0.179114 | 0.22453 | 0.21075 |
|  | $\pi_{4}$ | 1.75436 | 0.0156827 | 5.3347231 | 0.139908 | 0.25037 | 0.21295 |
|  | $\pi_{5}$ | 3.34804 | 0.5813534 | 5.8202566 | 0.259134 | 0.25545 | 0.21595 |





Figure 2 Prior distriburtions of the change point


Figure 3 Marginal posterior distributions under prior $\pi_{1}(t)$


Figure 4 Marginal posterior distributions under prior $\pi_{2}(t)$


Figure 5 Marginal posterior distributions under prior $\pi_{3}(t)$


Figure 6 Marginal posterior distributions under prior $\pi_{4}(t)$


Figure 7 Marginal posterior distributions under prior $\pi_{5}(t)$

## §5．Conclusion and Remarks

Beta－binomial distribution is selected as the prior of the change point，because beta－ binomial distribution with different parameters can well describe the prior information of the change point．And power that incorporate the historical data into the current study can well describe the historical information．According to the properties of $\chi^{2}$ distribution and Student＇s $t$ distribution，we get closed forms of the marginal posterior distributions of the change point，the regression coefficients and the common variance．Simulation shows higher performance or accuracy of the Bayesian method when the prior distribution of the change point follows the beta－binomial distribution．

Some further work can be done and under our investigation．
1．regression models with difference variances；
2．regression models with multiple change points；
3．regression models with changes occur as independent variables instead of time or observation，as mentioned in MacNeill and Mao（1995），Gallant（1977）and Pastor and Guallar（1998）；

4．generalized linear models with change points．

## References

［1］Chin Choy，J．H．and Broemeling，L．D．，Some Bayesian inferences for a changing linear model，Tech－ nometrics，22（1）（1980），71－78．
［2］Chen，J．and Gupta，A．K．，Testing and locating variance changepoints with application to stock prices，Journal of the American Statistical Association，92（438）（1997），739－747．
［3］Deya，D．K．and Purkayastha，S．，Bayesian approach to change point problems，Communications in Statistics－Theory and Methods，26（8）（1997），2035－2047．
［4］Ferreira，P．E．，A Bayesian analysis of a switching regression model：known number of regimes，Journal of the American Statistical Association，70（350）（1975），370－374．
［5］Gallant，A．R．，Testing a nonlinear regression specification：a nonregular case，Journal of the American Statistical Association， $\mathbf{7 2 ( 3 5 9 ) ( 1 9 7 7 ) , ~ 5 2 3 - 5 3 0 . ~}$
［6］Ibrahim，J．G．，Chen，M．H．and Sinha，D．，On optimality properties of the power prior，Journal of the American Statistical Association，98（461）（2003），204－213．
［7］MacNeill，I．B．and Mao，Y．，Change－point analysis for mortality and morbidity rate，In Applied Change Point Problems in Statistics（Editors：Sinha，B．，Rukhin，A．and Ahsanullah，M．），1995，37－ 55.
［8］Pastor，R．and Guallar，E．，Use of two－segmented logistic regression to estimate change－points in epidemiologic studies，American Journal of Epidemiology，148（7）（1998），631－642．
［9］Quandt，R．E．，The estimation of the parameters of a linear regression system obeying two separate regimes，Journal of the American Statistical Association，53（284）（1958），873－880．
［10］Zhou，H．and Liang，K．－Y．，On estimating the change point in generalized linear models，Institute of Mathematical Statistics Collections，1（2008），305－320．

## 线性模型变点问题的贝叶斯分析

汤银才 王平平<br>（华东师范大学金融与统计学院，上海，200241）

陈 惠<br>（上海西南位育中学，上海，200233）

本文主要讨论了变点的先验分布为beta－binomial分布和Ibrahim等（2003）提出的幂型先验的条件下，有一个变点的线性模型的贝叶斯统计推断问题，并且我们假定变点两边的观测值的方差是相等的。我们得到变点，回归系数，共同方差的后要分布的显示表达式。本论文不仅把Ferrira（1975）论文从变点先验分布服从离散均匀分布推广到了更好描述变点的形状的beta－binomial分布，而且进一步将变点的先验分布推广到包含的历史信息的幂型先验。当变点的先验分布为beta－binomial分布和幂型先验时，模拟结果显示了贝叶斯方法具有更高的准确性。

关键词：beta－binomial先验，幂型先验，变点，贝叶斯估计，线性模型。
学科分类号：O212．1．


[^0]:    ${ }^{*}$ The research was supported by Natural Science Foundation of China（11271136），Program of Shanghai Subject Chief Scientist（14XD1401600）and the 111 Project（B14019）．
    ＊Corresponding author，E－mail：yctang＠stat．ecnu．edu．cn．
    Received September 10，2014．Revised October 14， 2014.
    doi：10．3969／j．issn．1001－4268．2015．01．008

